Optimization of replica exchange molecular dynamics by fast mimicking

Hritz, J.; Oostenbrink, C.

published in
Journal of Chemical Physics
2007

DOI (link to publisher)
10.1063/1.2790427

document version
Publisher's PDF, also known as Version of record

Optimization of replica exchange molecular dynamics by fast mimicking

Jozef Hritz and Chris Oostenbrink[a)]
Leiden Amsterdam Center for Drug Research (LACDR), Division of Molecular Toxicology, Vrije Universiteit, Amsterdam, The Netherlands

(Received 8 August 2007; accepted 5 September 2007; published online 27 November 2007)

We present an approach to mimic replica exchange molecular dynamics simulations (REMD) on a microsecond time scale within a few minutes rather than the years, which would be required for real REMD. The speed of mimicked REMD makes it a useful tool for "testing" the efficiency of different settings for REMD and then to select those settings, that give the highest efficiency. We present an optimization approach with the example of Hamiltonian REMD using soft-core interactions on two model systems, GTP and 8-Br-GTP. The optimization process using REMD mimicking is very fast. Optimization of Hamiltonian-REMD settings of GTP in explicit water took us less than one week. In our study we focus not only on finding the optimal distances between neighboring replicas, but also on finding the proper placement of the highest level of softness. In addition we suggest different REMD simulation settings at this softness level. We allow several replicas to be simulated at the same Hamiltonian simultaneously and reduce the frequency of switching attempts between them. This approach allows for more efficient conversions from one stable conformation to the other. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2790427]

I. INTRODUCTION

Since the introduction of the replica exchange method (REM) using the Monte Carlo algorithm, replica exchange molecular dynamics (REMD) or parallel tempering simulations in the late 1990s,[1–5] there has been a steep increase in their popularity. In these simulations schemes the conformational sampling of a molecular system is greatly enhanced by connecting multiple simulations that are performed at different temperatures [temperature REMD (*T*-REMD)] or using different functional forms of the potential interaction energy [Hamiltonian REMD (*H*-REMD)].[6] By maintaining a detailed balance requirement when individual simulations are switched to a different temperature or Hamiltonian, the correct thermodynamic ensemble will be obtained for each of the simulation parameters. An intelligent choice of REMD settings allows for the swift generation of canonical ensembles of systems in which the potential energy barriers between stable conformations are too large to be crossed repeatedly in a normal molecular dynamics (MD) simulation. In this paper we will introduce an approach to efficiently optimize the settings for REMD simulation for systems with multiple stable conformations. Settings to optimize involve the number of simultaneous simulations (replicas) and the optimal simulation settings for each of these simulations (temperature, Hamiltonian).

Only researchers with access to extraordinary computational resources can afford a trial and error approach when searching for efficient REMD settings. Several studies describe approaches leading to efficient REMD simulations, mostly for *T*-REMDs. [Because one MD step requires the same amount of CPU time for any temperature or Hamiltonian, the allocation of replicas to CPUs can be trivial (one replica per CPU). However, in replica exchange Monte Carlo simulations, the average wall clock time to complete one Monte Carlo move varies with the temperature or the Hamiltonian, and the CPU allocation becomes an important issue.[7]] Many authors claim that the highest efficiency is obtained when the switching probability between neighboring replicas is constant at a value of approximately 20%.[8–13] This is still the most often used criterion in REMD applications despite the fact that in 2004 Trebst *et al.* presented the feedback-optimized parallel tempering approach.[14] It was shown that for the optimal temperature sets the switching probabilities between neighboring replicas are not constant but rather depend on the temperature.[14–16] A significant practical drawback of this approach, however, is the simulation time required to obtain the optimized settings. Especially for biomolecular simulations, this hampers the practical applicability of the method. The application of feedback-optimized parallel tempering for the 36-residue villin headpiece subdomain HP-36 required REMD simulations that covered 400 000 switching trials, which takes many years of CPU time.[16] It is probably for this reason that less applications of the method have been described.

For this reason we have developed a set of efficient tools for optimizing the settings of REMD (both *T*- and *H*-REMD) simulations for systems, of which multiple stable conformations are known. By generating the appropriate conformational ensemble of the system REMD gives insight into the relative populations of stable conformations. We present the practical application of the proposed optimization scheme for two biologically relevant systems: GTP and 8-Br-GTP (Fig. 1). GTP is an important component in, e.g., cellular signaling, while 8-Br-GTP is considered as promising anti-

[a)]Electronic mail: c.oostenbrink@few.vu.nl

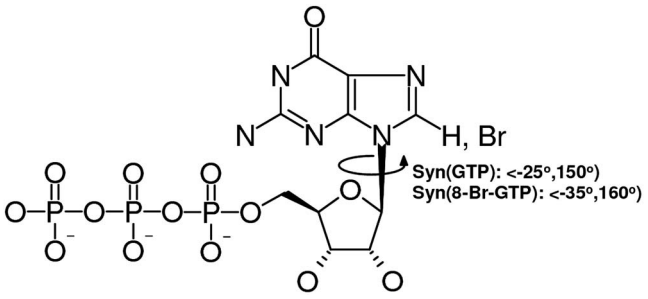

FIG. 1. Structure of GTP and 8-Br-GTP in syn conformation. Conformational transitions between the syn and anti states occur by rotation around the glycosidic bond (indicated).

bacterial agent. It inhibits FtsZ polymerization but does not affect tubulin polymerization.[17] GTP and 8-Br-GTP both have two stable orientations of the base toward the sugar: anti and syn (Fig. 1). Nuclear magnetic resonance (NMR) studies show that while GTP prefers to be in the anti conformation (with an estimated population of ~70% from NMR experiments), 8-Br-GTP shows preference for the syn conformation (population ~90%). Our MD study shows that there is a high energy barrier between the anti and syn conformations for both molecules. In two MD simulations of GTP starting from anti and syn conformation (each of 20 ns) only one single transition from syn to anti was observed (after ~1 ns) while no anti→syn transition was seen at all. Analogous simulations of 8-Br-GTP did not reveal any transition indicating an even higher potential energy barrier between the anti and syn conformations.[18] In these cases REMD can be used to enhance the conformational sampling and obtain the correct conformational ensemble with the proper populations of syn and anti conformations. From this ensemble, the free energy difference between the two states as well as a variety of molecular properties can be calculated.

II. METHODS

A. MD settings

All MD simulations were conducted using the GROMOS05 MD simulation package running on a linux cluster.[19] All bonds were constrained, using the SHAKE algorithm[20] with a relative geometric accuracy of 10^{-4}, allowing for a time step of 2 fs used in the leapfrog integration scheme.[21] Periodic boundary conditions, with a truncated octahedral box, were applied. After the steepest descent minimization to remove bad contacts between molecules, the initial velocities were randomly assigned from a Maxwell–Boltzmann distribution at 298 K, according to the atomic masses. The temperature was controlled using a weak coupling to a bath of 298 K with a time constant of 0.1 ps.[22] The solute molecules (GTP or 8-Br-GTP) and solvent (i.e., explicit water molecules and 3 Na^+ counterions) were independently coupled to the heat bath. The pressure was controlled using isotropic weak coupling to atmospheric pressure with a time constant 0.5 ps.[22] Van der Waals and electrostatic interactions were calculated using a triple range cutoff scheme. Interactions within a short-range cutoff of 0.8 nm were calculated every time step from a pair list that was generated every five steps.

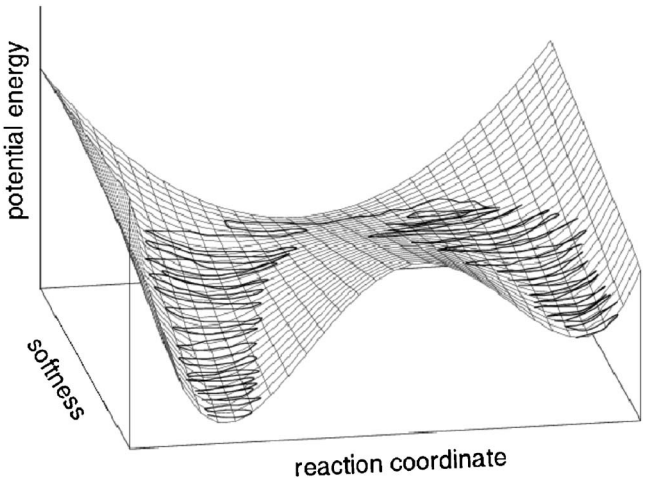

FIG. 2. Schematic figure of the energy landscape profile as function of the softness of nonbonded interactions. In the presented *H*-REMD using soft-core interactions, the replicas at higher levels of softness have a higher probability for the conformational transition between two stable conformational states. Still, this transition requires some time.

At these time points, interactions between 0.8 and 1.4 nm were also calculated and kept constant between updates. A reaction-field contribution was added to the electrostatic interactions and forces to account for a homogeneous medium outside the long-range cutoff, using the relative permittivity of SCP water (61).[23] All interaction energies were calculated according to the GROMOS force field, parameter set 53A6.[24] Force field parameters used for GTP and 8-Br-GTP are listed in the supplementary material of Ref. 18.

In this work we will focus on the optimization of a *H*-REMD approach in which the modification of the Hamiltonian consists of a softening of selected interactions. For this reason we have used the following form for Van der Waals and electrostatic soft-core interactions as a function of the interatomic distance r_{ij}:[25]

$$E_\lambda^{\mathrm{vdw}}(r_{ij}) = \left(\frac{C_{12}}{A_\lambda + r_{ij}^6} - C_6\right)\frac{1}{A_\lambda + r_{ij}^6}, \tag{1}$$

$$E_\lambda^{\mathrm{el}}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon}\frac{1}{\sqrt{B_\lambda + r_{ij}^2}}, \tag{2}$$

where $A_\lambda = \alpha_{\mathrm{vdw}}(C_{12}/C_6)\lambda^2$ and $B_\lambda = \alpha_{\mathrm{el}}\lambda^2$. C_{12} and C_6 are the Lennard–Jones parameters, q_i and q_j are the partial charges of particles i and j, and α_{vdw} and α_{el} are the softness parameters that can be set for every pair of atoms. In the current study we used in all simulations $\alpha_{\mathrm{vdw}} = \alpha_{\mathrm{el}} = 1$ and the softness of the interactions was controlled through the λ parameter. It can be seen that at longer distances the soft-core interaction approximates the interaction for normal atoms and that they differ mostly at short distances between atoms. Potential energy barriers are mostly the result of short-ranged repulsion between atoms, which can strongly be reduced at higher levels of softness (Fig. 2). In this study, only interactions between the nucleotide base and sugar are treated using the soft-core interaction.

The systems were first equilibrated for 100 ps MD at constant pressure, where position restraints were applied on

heavy atoms of GTP/8-Br-GTP, after which another 100 ps of equilibration at constant pressure without any restraints were carried out. Finally, the systems were simulated for 1 ns at different λ values, where coordinates of the whole system were recorded every 1 ps. Ten independent 1 ns MD simulations at different levels of softness were performed for GTP ($\lambda=0.0,0.05,\ldots,0.4,0.45$) and 15 for 8-Br-GTP ($\lambda=0.0,0.05,\ldots,0.65,0.7$).

B. REMD

In a REMD simulation, a number of noninteracting MD runs (called replicas) are simulated at different conditions. Let us label replicas as $0,1,\ldots,n$. After a given time (elementary period, t^{elem}), an exchange between two neighboring replicas is attempted, followed by another set of independent MD simulations. In the GROMOS05 implementation switches are first attempted between pairs $0\leftrightarrow1,2\leftrightarrow3,\ldots$ (type I of replica exchange trials) and after the next t^{elem} of MD simulations switches are subsequently attempted between pairs $1\leftrightarrow2,3\leftrightarrow4,\ldots$ (type II of replica exchange trials). This means that the effective time between switching attempts of the same type is twice the elementary period. In our study we used $t^{\text{elem}}=2.5$ ps leading to an effective period of 5 ps between identical switching attempts.

In contrast to *T*-REMD, where the temperature is increased to facilitate the crossing of high energy barriers, we are modifying the Hamiltonian within *H*-REMD by making use of the soft-core interactions, Eqs. (1) and (2). This will lead to a decrease of the potential energy barrier between conformations and thus also facilitate the transition from one potential energy minimum into the other (see Fig. 2). The proper ensemble of conformations at every value of λ can be obtained by applying the Metropolis criterion for the exchange probability w_{λ_i,λ_j} between neighboring replicas running at Hamiltonians corresponding to the parameters λ_i and λ_j,

$$w_{\lambda_i,\lambda_j}=\min[1,\exp(-\Delta_{\lambda_i,\lambda_j})], \tag{3}$$

where

$$\Delta_{\lambda_i,\lambda_j}=\beta[E_{\lambda_j}(q_{\lambda_i})-E_{\lambda_i}(q_{\lambda_i})+E_{\lambda_i}(q_{\lambda_j})-E_{\lambda_j}(q_{\lambda_j})]. \tag{4}$$

The coordinates q_{λ_i} represent a conformation that was obtained from a simulation at the Hamiltonian corresponding to parameter λ_i and E_{λ_j} is the potential energy calculated according to the Hamiltonian corresponding to the parameter λ_j [Eqs. (1) and (2)]. $\beta=1/k_BT$ with k_B as the Boltzmann constant and T as the absolute temperature.

C. Concept of double/multiple replicas at the highest lambda/temperature

The elementary period, t^{elem}, in REMD should be long enough to relax the energy before the next switching attempt. Otherwise one observes many "reswitches" at the next switching attempt of the same type after a replica switch with low probability. Of course one can increase t^{elem}, but then the overall REMD efficiency decreases with decreasing number of replica switches. Therefore, a reasonable balance will need to be struck.

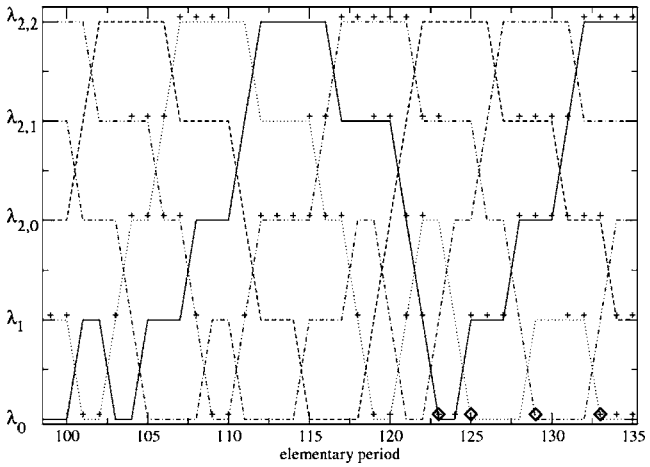

FIG. 3. Schematic example of real REMD with three replicas at $\lambda_{\max}$ ($\lambda_{2,0}=\lambda_{2,1}=\lambda_{2,2}=\lambda_{\max}$) and one middle λ_1 illustrating global conformational transitions, indicated by a diamond symbol. Crosses indicate the presence of syn conformation. Switches between replicas at $\lambda_{2,0}$ and $\lambda_{2,1}$ are attempted only at times, t, which are a multiple of 5 t^{elem} and between $\lambda_{2,1}$ and $\lambda_{2,2}$ if $t+t^{\text{elem}}$ is a multiple of 5 t^{elem}. It ensures that a replica that switches from $\lambda_{2,0}$ to $\lambda_{2,1}$ will spend the required time, $t^{\text{total}}_{\lambda_{\max}}=10t^{\text{elem}}$ at $\lambda_{2,1}$ and $\lambda_{2,2}$ altogether.

However, there are more relaxation processes that play a role. Even at values of λ where the barrier has been removed or at temperatures where it is readily crossed, the system still needs time to move from one conformation to the other. (Note: Even for a very small conformational barrier the average transition time is significantly longer than times generally used in REMD between switching trials. This is mainly true for transitions to the less favorable conformational states.) The occurrence of syn states for a set of simulations starting at anti under the conditions of the highest λ value or temperature and the occurrence of anti states for simulations starting from a syn conformation as a function of time, will typically be represented by a sigmoidal curve. We define the transition time $t^{\text{transit}}_{\lambda_{\max}}$ as the time at which this sigmoidal curve reaches saturation. The height of the curve allows us to estimate the relative populations of syn and anti at this λ value and from that the transition probabilities at times larger than $t^{\text{transit}}_{\lambda_{\max}}$, $P^{\text{anti}\rightarrow\text{syn}}_{\lambda_{\max}}(t^{\text{transit}}_{\lambda_{\max}})$, and $P^{\text{syn}\rightarrow\text{anti}}_{\lambda_{\max}}(t^{\text{transit}}_{\lambda_{\max}})$.

The conformational transition time, $t^{\text{transit}}_{\lambda_{\max}}$, is for the majority of systems much longer than t^{elem}. If t^{elem} would be extended to $t^{\text{transit}}_{\lambda_{\max}}$, then the efficiency gain of REMD becomes very low. In order to combine frequent switching trials with long enough relaxation times to allow for anti↔syn conformational transitions to occur at $\lambda_{\max}$ we introduce a new scheme, called degenerated $\lambda_{\max}$ scheme, involving multiple (n) replicas at $\lambda_{\max}$ (or at the highest temperature for *T*-REMD).

This can be done by "degenerating" $\lambda_{\max}$ into $\lambda_{\max,0},\lambda_{\max,1},\ldots,\lambda_{\max,n-1}$. In the switching scheme all λ values are ordered as $\lambda_0,\lambda_1,\ldots,\lambda_{\max-1}$, $\lambda_{\max,0},\lambda_{\max,1},\ldots,\lambda_{\max,n-1}$, and only switches are attempted between neighboring λ values. The switching frequency between replicas at $\lambda_{\max}$ is reduced by only allowing switching attempts after a given multiple of the elementary period (Fig. 3). We define the time $t^{\text{total}}_{\lambda_{\max}}$ as the total simulation time be-

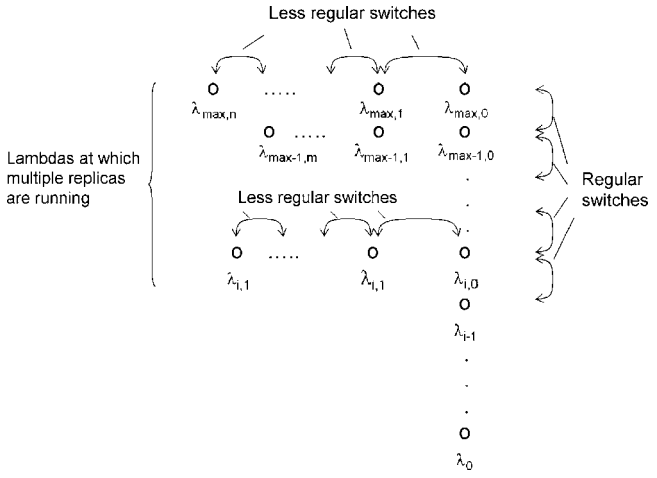

FIG. 4. Schematic figure showing the application of degenerate λ's. By limiting the number of switching attempts between replicas at the same λ value, the systems are allowed more time to relax toward different conformations at this value of λ.

tween the switch $\lambda_{max,0} \rightarrow \lambda_{max,1}$ until the return switch $\lambda_{max,1} \rightarrow \lambda_{max,0}$. Note that the switching probability between replicas at λ_{max} equals 1 because they correspond to the same Hamiltonian and Δ in Eq. (4) becomes 0 by definition. The procedure to find the optimal $t^{total}_{\lambda_{max}}$ is described in Sec. II D 2. The switching trial frequency between all other pairs of replicas is much higher (with an elementary period of 2.5 or 5 ps between two switching trials of the same type), which maintains an efficient diffusion of replicas between the lowest and highest λ value.

An increasing number of replicas at λ_{max} increases the convergence of conformational populations at λ_{max}, which subsequently leads to a faster convergence of conformational populations at $\lambda_{max-1}, \lambda_{max-2}, \ldots$ and finally to the faster convergence of populations over the whole REMD system. The REMD efficiency gain when using more replicas at λ_{max} comes from the fact that multiple conformational transition "attempts" are performed in parallel.

With n replicas at λ_{max}, a replica that has had sufficient time to show conformational transition becomes available at $\lambda_{max,0}$ with a period of $(1/n-1)t^{total}_{\lambda_{max}}$. This allows for values of n up to $n_{max}=t^{total}_{\lambda_{max}}/t^{elem}+1$, although the efficiency is limited in this case because replicas tend to stay for several periods of $t^{total}_{\lambda_{max}}$ at λ_{max} rather than get an opportunity to switch down to λ_{max-1}. For this reason, the maximal overall efficiency is obtained at $n \approx w_{\lambda_{max-1},\lambda_{max,0}} n_{max}$. Especially for complex systems the required time $t^{total}_{\lambda_{max}}$ at λ_{max} can be high, so a higher number of replicas at λ_{max} can significantly increase the overall REMD efficiency, almost by a factor $(n-1)$. For the examples described in this work, a one-dimensional setup with multiple replicas at λ_{max} seems sufficient, but the concept can easily be extended to having a degenerate set of replicas at a certain λ value or temperature by going to a second or third dimension of λ or T. See Fig. 4.

D. Searching for the optimal REMD settings

The efficiency of REMD to sample conformational space depends mainly on the efficiency of conformational transitions at λ_{max} and an efficient diffusion of replicas between the lowest and highest λ values. In this section, we describe the tools that are used to optimize the choice of λ values and the number of replicas at λ_{max}.

Section II D 1 describes the algorithm we propose for the fast mimicking of REMD without performing actual MD. This algorithm requires knowledge of the optimal value of λ_{max}, the optimal conformational conversion time at λ_{max}, $t^{total}_{\lambda_{max}}$, as well as estimates of the saturated conformational transition probabilities, $P^{anti\rightarrow syn}_{\lambda_{max}}$, $P^{syn\rightarrow anti}_{\lambda_{max}}$. In addition, the algorithm makes use of switching probabilities between replicas at different λ values. Section II D 2 describes the selection of λ_{max}, the calculation of $t^{total}_{\lambda_{max}}$, and conversion probabilities. Section II D 3 describes how switching probabilities are sampled from precalculated probability distributions and Sec. II D 4 finally describes the approach we take to use the mimicking algorithm to obtain the most efficient settings for REMD simulations.

1. Mimicked REMD

Mimicked REMD assumes a number of discreet stable conformational states exists, labeled, $c=0,1,\ldots,N_c-1$ (in our case anti and syn). In the present case these conformations are described as anti and syn. Instead of performing MD we estimate the probability of a conformational transition as well as the REMD switching probabilities and subsequently propagate the conformational states through time. This requires knowledge of the probability distributions of switching probabilities $\{w^{c_k,c_l}_{\lambda_i,\lambda_j}\}$ for all neighboring pairs of λ values: λ_i, λ_j as well as approximate probabilities of the conformational changes at each λ value from one stable conformational region to the other ones. We would like to note that if precise conformational transition probabilities would be known then no REMD mimicking is needed, as one could easily derive the individual conformational populations as well as the free energy difference between the stable conformations. The merit of the current approach comes from the fact that the transition probabilities can be estimated for λ_{max} fairly easily, while they are ~0 for all other λ values.

As explained in Sec. II C the transition between conformational states is a dynamic process in which the conformational transitions depend sigmoidally on time. From these curves at λ_{max}, one can estimate the probability of the transition $c_i \rightarrow c_j$, after the (prolonged) residence time, $P^{c_i\rightarrow c_j}_{\lambda_{max}}(t^{transit}_{\lambda_{max}})$. Because λ_{max} will be selected such that the corresponding $t^{total}_{\lambda_{max}}$ is shortest (see Sec. II D 2) and because for all other replicas the time between switching attempts $(2t^{elem})$ is much shorter than the corresponding t^{total}_{λ}, it is reasonable to assume that the probabilities of conformational transitions at any λ value other than λ_{max} approach zero $[P^{c_i\rightarrow c_j}_{\lambda\neq\lambda_{max}}(2t^{elem})=0]$.

In our REMD mimicking algorithm we begin by assigning starting conformational states to all replicas. This can be done either randomly, sampled from the correct ensemble, or biased toward one of the states.

Subsequently four steps make up the main cycle of our algorithm:

(1) Intralambda conformational transitions. We predict if the conformational state changes during the elementary period (in our case $t^{\text{elem}}=2.5$ ps) at each λ value by the given probabilities. As described above all conformational transition probabilities are set to zero for all λ values, except for $\lambda_{\max}$, where $P^{c_i\to c_j}_{\lambda_{\max}}(t^{\text{total}}_{\lambda_{\max}}) = P^{c_i\to c_j}_{\lambda_{\max}}(t^{\text{transit}}_{\lambda_{\max}})$ is nonzero once per number of cycles corresponding to the $t^{\text{total}}_{\lambda_{\max}}$ spent at $\lambda_{\max,1},\ldots,\lambda_{\max,n-1}$.
(2) Replica exchanges of type I (between $0\leftrightarrow 1, 2\leftrightarrow 3,\ldots$). According to the actual conformational state at each λ value we take the corresponding probability distributions of switching probabilities $\{w^{c_k,c_l}_{\lambda_i,\lambda_j}\}$ for all neighboring λ pairs of type I. A switching probability is randomly chosen from the distribution of switching probabilities $\{w^{c_k,c_l}_{\lambda_i,\lambda_j}\}$ and a second random number determines if the switch occurs.
(3) Intralambda conformational transitions. The same as the first step.
(4) Replica exchanges of type II (between $1\leftrightarrow 2$, $3\leftrightarrow 4,\ldots$). The same as the second step but now for the λ pairs of type II.

The length of the whole cycle corresponds to the double of elementary switching attempt period (in our case 2×2.5 ps$=5$ ps). Sampling of a half million cycles (corresponding to 10^6 t^{elem} of real REMD) takes typically 15 min (depending on the exact number of replicas) using an unoptimized python script.

2. Selection of $\lambda_{\max}$

The average conformational transition time at the unperturbed Hamiltonian ($\lambda=0.0$), $t^{\text{transit}}_{0.0}$ is by far too long to sample sufficient transitions reversibly. In *H*-REMD, the Hamiltonian is parameterized such that the potential energy barrier associated to the conformational transition is reduced for higher values of λ, thereby also reducing $t^{\text{transit}}_{\lambda_i}$.

To obtain the optimal value of $\lambda_{\max}$, we perform ten short MD simulations (200 ps) starting from anti and 10 short MD runs starting from syn conformation at different λ values (for GTP $\lambda=0.4;0.45;0.5$ and for 8-Br-GTP $\lambda=0.6;0.65;0.7;0.8;0.9$). (Note: $\lambda_{\max}$–values lower than 0.4 respectively. 0.6 were not considered based on the results of 1 ns runs at different λ values used for the calculation of switching probability distributions as described in the next paragraph.)

Our aim is to find the value of $t^{\text{total}}_{\lambda_{\max}}$ for which we would get the highest number of conformational transitions at $\lambda_{\max}$ during a given length of REMD simulation (see also Sec. II C). While the time dependency of the number of simulations where the conformation has changed with respect to the initial conformation has an S-curve profile, it is clear that the most efficient $t^{\text{total}}_{\lambda_{\max}}$ will be in the interval between the midpoint of the S curve and the saturated region. Quantitatively $t^{\text{total}}_{\lambda_{\max}}$ is obtained as the time corresponding to the maximum of the number of conformational changes divided by time. Because these maxima occur at different times for the anti $\to$ syn transition and the syn $\to$ anti transitions, and because we need to have a large enough number of both types of transitions, we select the longer time of both transitions at one $\lambda_{\max}$ value. We select the value of $\lambda_{\max}$ for which $t^{\text{total}}_{\lambda_{\max}}$ is shortest and where enough transitions occur in both directions. The set of ten MD runs starting from different conformations is subsequently prolonged for the selected $\lambda_{\max}$, in order to refine the converged values of the conformational conversion probabilities $P^{\text{anti}\to\text{syn}}_{\lambda_{\max}}(t^{\text{transit}}_{\lambda_{\max}})$ and $P^{\text{syn}\to\text{anti}}_{\lambda_{\max}}(t^{\text{transit}}_{\lambda_{\max}})$. Because the exact shape of the S curve is not properly converged from only ten simulations, it is not wise to directly read $P^{\text{transit}}_{\lambda_{\max}}(t^{\text{total}}_{\lambda_{\max}})$ from these curves. Rather, we set $P^{\text{transit}}_{\lambda_{\max}}(t^{\text{total}}_{\lambda_{\max}}) = P^{\text{transit}}_{\lambda_{\max}}(t^{\text{transit}}_{\lambda_{\max}})$, a value which converges also for a limited number of simulations. Using $P^{\text{transit}}_{\lambda_{\max}}(t^{\text{transit}}_{\lambda_{\max}})$ also ensures the generating the proper conformational populations at $\lambda_{\max}$ within the mimicked REMD.

3. Generation of probability distributions of switching probabilities

Let us consider the system, which has several different stable conformational sates (syn, anti), for which standard MD simulations do not produce enough transitions due to the conformational barriers between these states. We assume that the conformational space within each of these stable regions is sampled "reasonably" well by a relatively short MD simulation (~ 1 ns). The description below follows the schematic representation in Fig. 5.

To obtain the probability distributions of the switching probabilities needed for the REMD mimicking we performed a set of 1 ns MD simulations starting from different stable conformers at different values of λ (alternatively, one could use different temperatures) between 0.0 and $\lambda_{\max}$ usually at increments of 0.05 (in principle one can use also REMD at these λ values for this purpose). Every 1 ps we store one conformational configuration frame leading to 1000 different configurations. It is possible that during these MD simulations conformational transitions occur. For this reason, we collect conformations belonging to the same stable conformational region *c* and the same value of λ at which the simulation was performed into one set of conformations

$$\{q^c_\lambda\}, \lambda = 0, 0.05, 0.1, \ldots, \lambda_{\max};$$

$$c = 0, 1, \ldots, N_c - 1.$$

For every combination of λ and *c* we obtain about 1000 different structures. The *r*th structural frame from this "trajectory" is expressed as $q^c_\lambda(t_r)$. For each of these structures we calculate its energy not only using the Hamiltonian at which it was simulated (represented by λ_i), but also the energy according to the Hamiltonian corresponding to all other λ values (λ_j), and expressed as: $E_{\lambda_j}(q^c_{\lambda_i}(t_r))$.

For each pair of configurations $q^{c_k}_{\lambda_i}(t_r)$; $q^{c_l}_{\lambda_j}(t_s)$ we calculate switching energy difference

$$\Delta^{c_k,c_l}_{\lambda_i,\lambda_j}(t_r,t_s) = \beta[E_{\lambda_j}(q^{c_k}_{\lambda_i}(t_r)) - E_{\lambda_i}(q^{c_k}_{\lambda_i}(t_r)) + E_{\lambda_i}(q^{c_l}_{\lambda_j}(t_s)) - E_{\lambda_j}(q^{c_l}_{\lambda_j}(t_s))] \quad (5)$$

and the corresponding switching probability

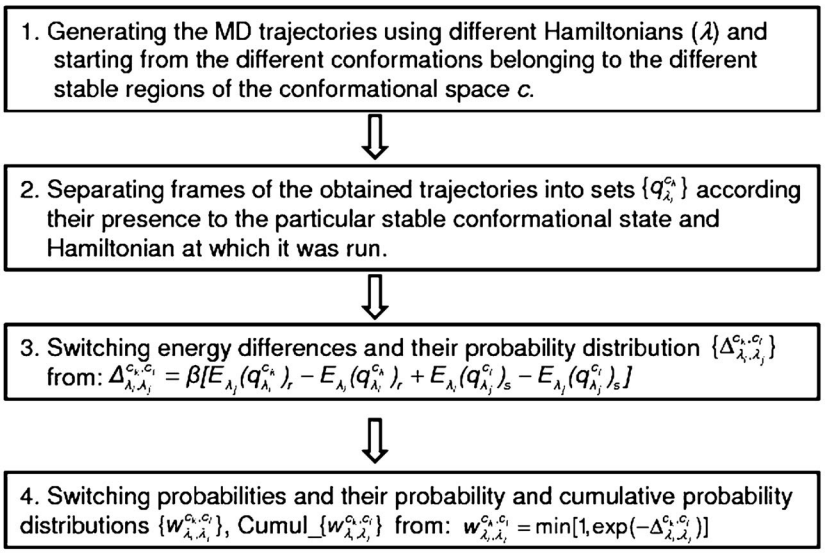

FIG. 5. Scheme describing the steps for obtaining the probability distribution of the switching probabilities.

$$w_{\lambda_i,\lambda_j}^{c_k,c_l}(t_r,t_s) = \min[1,\exp(-\Delta_{\lambda_i,\lambda_j}^{c_k,c_l}(t_r,t_s))], \quad (6)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the absolute temperature.

The distributions of $\Delta_{\lambda_i,\lambda_j}^{c_k,c_l}$ and $w_{\lambda_i,\lambda_j}^{c_k,c_l}$ values are thus calculated from altogether $\sim 10^6$ values, which are marked as $\{\Delta_{\lambda_i,\lambda_j}^{c_k,c_l}\}$ and $w_{0.0,0.3}^{\text{syn,anti}}$ after normalization. This gives us the probability distribution for a randomly chosen pair of configurations, $q_{\lambda_i}^{c_k}$, $q_{\lambda_j}^{c_l}$ with the switching probability between them given by $w_{\lambda_i,\lambda_j}^{c_k,c_l}$. For the REMD mimicking we found it convenient to produce Cum_$\{w_{\lambda_i,\lambda_j}^{c_k,c_l}\}$, which contains the cumulants of the probability distribution of switching probabilities.

4. Searching algorithm for the parameters of the most efficient REMD

500 000 cycles of REMD mimicking (corresponding to 10^6 t^{elem} of real REMD, in our case 2.5 μs) takes ~ 15 min which allows us to perform "REMD" for hundreds of different combinations of λ values and numbers of replicas at $\lambda_{\max}$ producing time sequences of the occurrence of conformational states at different $w_{0.0,0.3}^{\text{syn,anti}} = \min[1,\exp(-\Delta_{0.0,0.3}^{\text{syn,anti}})]$ values. From these one can decide for different criteria from which to prefer one set of parameters over the other.

For cases where different conformational states can be defined (e.g., the anti and syn states), we propose to measure efficiency of REMD, through the "number of global conformational transitions," N_{gct}. We count the number of conformational transitions for each particular replica at the lowest λ value, λ_0.

The schematic REMD example shown in Fig. 3, contains three replicas at $\lambda_{\max}$ ($\lambda_{2,0} = \lambda_{2,1} = \lambda_{2,2} = \lambda_{\max}$) and one middle value of lambda, λ_1. The replica represented by a solid curve, is in the anti conformation at λ_0 after 103 t^{elem} and then exchanges to higher λ values. At $\lambda_{2,1}$ it makes the transition to the syn conformation (after 119 t^{elem}) and subsequently exchanges with replicas at lower λ values until it finally returns to λ_0, but now in the different conformational state (syn) as before. The combination of replica exchanges and a conformational transition is counted as one global anti→syn transition. From this moment onwards, we will monitor for the next syn→anti global conformational transition (for the same replica), etc. From this description one can see that for a global conformational transition to occur it is not necessary for the replica to reach $\lambda_{\max}$, because the conformational transition can occur also at lower λ values (see, e.g., in Fig. 3 the replica represented by a dotted line. At 131 t^{elem} an anti→syn transition occurs at λ_1 leading to the global conformational transition at 133 t^{elem}) or even at λ_0 within real REMD. (This is not the case for the mimicked REMD, where conformational transition can occur only at $\lambda_{\max}$.) On the other hand, the replica represented by the dotted curve is in a syn conformation at λ_0 at 102 t^{elem}, then went up to $\lambda_{\max}$ where it changes its conformation to anti and later back to syn again. When it exchanges back to λ_0 at 119 t^{elem}, this event is not counted as a global conformational transition, because the conformations at 102 t^{elem} and 119 t^{elem} are the same. In the case of MD (as a special case of REMD with only one replica at λ_0) N_{gct} is equal to the number of conformational transitions within one MD run. This allows for a direct comparison of the efficiency between very different REMDs or MDs performed under different conditions.

Our aim is to find the REMD settings which give us the maximum value of N_{gct} per CPU. (In the following, we assume that in real REMD simulations every replica is assigned one CPU.) Having decided on the optimal value of $\lambda_{\max}$ and $t_{\lambda_{\max}}^{\text{total}}$ we vary:

(1) the number and placement of middle λ's (between 0.0 and $\lambda_{\max}$) and
(2) the number of replicas at $\lambda_{\max}$.

Whether it is efficient to "invest" into more replicas at $\lambda_{\max}$ or not depends on the ratio of $t_{\lambda_{\max}}^{\text{total}}$ relative to the average time needed for global conformational transitions to occur. If more time is needed for a conformational change at $\lambda_{\max}$, it becomes more likely that an increased number of replicas at $\lambda_{\max}$ improves the overall efficiency per CPU.

In cases where only a few replicas are needed one can test various settings in a systematic manner and select the optimal combination. In spite of the efficiency of mimicked REMD, a larger number of required replicas will make the optimization problem more complex. We propose three useful approaches:

(1) The number and placement of λ values between 0.0 and λ_{max} (middle λ's) and the number of replicas at λ_{max} can be optimized independently. Middle λ's are essential for the diffusion of replicas between 0.0 and λ_{max}, which will be the same for any number of replicas at λ_{max}. The optimal setting of middle λ's will therefore be independent of the number of replicas at λ_{max}. However, the gain in efficiency due to optimization of the middle λ's is more pronounced if the conformational change at λ_{max} is not the bottleneck of the simulation. Therefore, we propose to search for the optimal settings with a high number of replicas at λ_{max}, initially 5 in the case of 8-Br-GTP. After this initial optimization, the number of optimal replicas at λ_{max} should be obtained after which one should check if having one more middle λ does not improve results even further. Once the optimal number of replicas at λ_{max} has been obtained for an optimal set of m middle λ values, this $n^{opt}(m)$ can be used to reduce the searching possibilities for different numbers of middle λ values by taking into account the following inequalities: $n^{opt}(k>m) \geq n^{opt}(m)$ and $n^{opt}(k<m) \leq n^{opt}(m)$.

(2) Searching for the optimal settings of middle λ's. When expanding or reducing the number of middle λ's in the optimization process, we do not need to consider the complete range (0.0, λ_{max}) for the new λ values. If we denote the optimal λ values in a scheme with n middle λ's as ${}^{n}\lambda_i^{opt}$ (${}^{n}\lambda_0^{opt}=0.0$; ${}^{n}\lambda_{n+1}^{opt}=\lambda_{max}$), then it is most likely that in a scheme with $n+1$ middle λ's, the optimal λ values are in the following range: ${}^{n+1}\lambda_i^{opt} \in \langle {}^{n}\lambda_{i-1}^{opt}, {}^{n}\lambda_i^{opt} \rangle$. Similarly, one can generally write that for a reduction of the number of middle λ's, the optimal λ values are restricted to ${}^{n}\lambda_i^{opt} \in \langle {}^{n+1}\lambda_i^{opt}, {}^{n+1}\lambda_{i+1}^{opt} \rangle$.

(3) In many cases, already a short simulation indicates if a certain set of settings is promising or not. In an approach we call "continuous filtering" we disregard possible settings on the fly, thereby only spending computer time on the most promising settings.

We perform a mimicked REMD simulation for all potentially relevant REMD settings for 30 000 cycles. We subsequently disregard all settings that show less than 10% of the (at that point) maximum value of N_{gct}. After 40 000 cycles this threshold is increased to 20% and by another 10% every 10 000 cycles until after 100 000 cycles only those settings are kept that lead to 80% of the maximum value of N_{gct}. We subsequently refine the search by increasing the threshold by 2.5% every 100 000 cycles and select the optimal setting after 500 000 cycles.

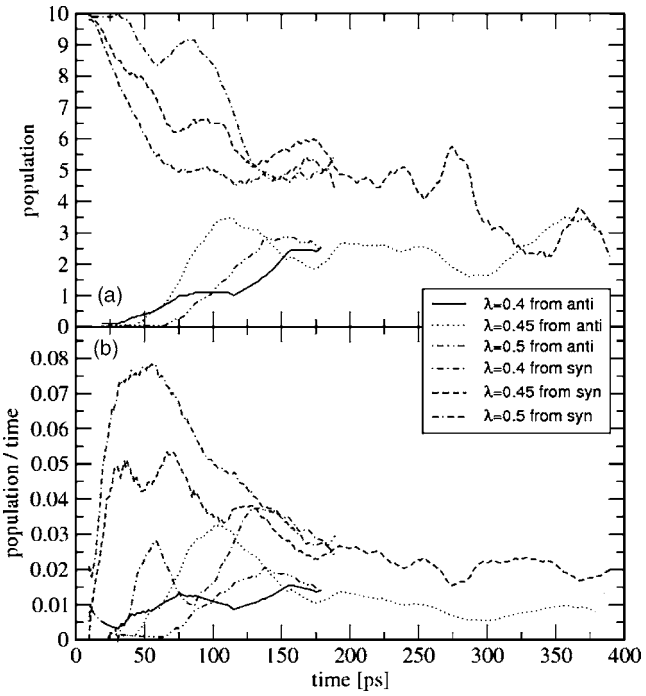

FIG. 6. (a) Number of syn conformations observed in a set of ten MD simulations of GTP as function of time at different λ values starting from ten different anti and syn conformations. Runs for λ=0.45 were prolonged up to 400 ps in order to reach convergence. (b) Population of converted conformations per time (syn→anti or anti→syn). For clarity, running averages over 20 time points have been taken.

III. RESULTS

Both GTP and 8-Br-GTP have two stable conformations: anti and syn, depending on the dihedral angle around the glycosidic bond (Fig. 1). Based on the distributions of this dihedral angle we define for GTP to be in the syn conformation, when its dihedral angle around the glycosidic bond is within the interval $<-25°, 150°)$, and otherwise to be in the anti conformation. For 8-Br-GTP the syn conformational interval is $<-35°, 160°)$.[18]

A. Selection of λ_{max}

MD simulations of GTP in explicit water were performed for 200 ps at different values of λ (0.4; 0.45; 0.5). Ten simulations started from anti and ten simulations started from syn conformations. During the MD runs conformational transitions occur.

Figure 6(a) show the number of simulations in which GTP is in a syn conformation as function of time, when starting from anti and when starting from syn conformation. We obtained $t_{\lambda_{max}}^{total}$ as the larger of two times corresponding to the maximas of the time dependence of the number of changed conformations divided by total time. From Fig. 6(b) follows: $t_{0.4}^{total}=155$ ps, $t_{0.45}^{total}=100$ ps, and $t_{0.5}^{total}=135$ ps. As the optimal value of λ_{max} we choose 0.45 because it gives the shortest $t_{\lambda_{max}}^{total}$ (100 ps) together with a sufficient number of conformational transitions after this time. Another advantage of $\lambda_{max}=0.45$ over 0.5 is the more efficient diffusion of replicas between λ_0 and λ_{max}. For $\lambda_{max}=0.45$, the set of ten MD simulations starting from syn and anti was prolonged leading

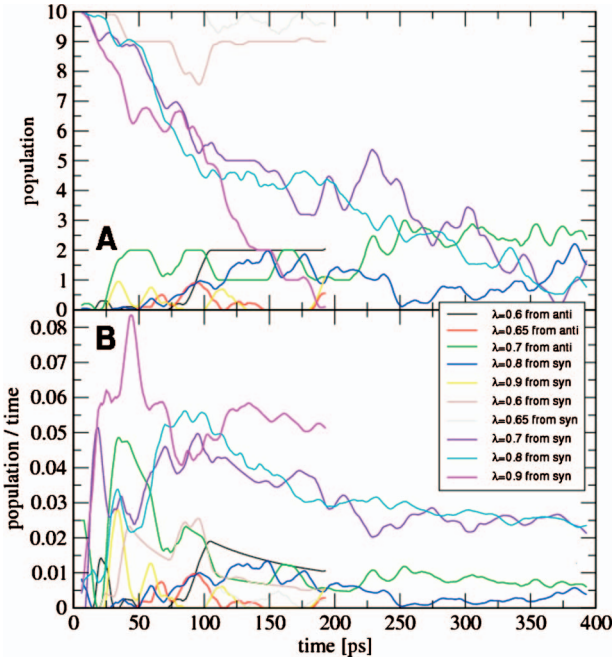

FIG. 7. (Color) (a) Number of syn conformations observed in a set of ten MD simulations of 8-Br-GTP as function of time at different λ values starting from ten different anti and syn conformations. Runs for λ=0.7 and λ=0.8 were prolonged up to 400 ps in order to reach convergence. (b) Population of converted conformations per time (syn→anti or anti→syn). For clarity, running averages over 20 time points have been taken.

to estimate of the converged conformational transition probabilities $P^{\text{anti}\to\text{syn}}_{\lambda=0.45}=0.3$ and $P^{\text{syn}\to\text{anti}}_{\lambda=0.45}=0.7$ [Fig. 6(a)].

The same kind of analysis was performed for 8-Br-GTP using λ=0.6;0.65;0.7;0.8;0.9 and leading to the population time dependences shown in Figs. 7(a) and 7(b). We considered as $\lambda_{\max}$ candidates $\lambda_{\max}=0.7$ and $\lambda_{\max}=0.8$ for which the prolonged MD simulations were performed. Based on these we decided to take $\lambda_{\max}=0.7$ for 8-Br-GTP with the corresponding $t^{\text{total}}_{\lambda_{\max}}=100$ ps and saturated transition probabilities $P^{\text{anti}\to\text{syn}}_{\lambda=0.7}=0.2$ and $P^{\text{syn}\to\text{anti}}_{\lambda=0.7}=0.8$. For λ=0.9 the transition syn→anti is quite efficient. However, the conversion probability from anti to syn has become very low which means that these transitions would occur only very rarely at λ=0.9.

B. Generation of probability distribution of switching probabilities

To obtain the probability distributions of switching probabilities we performed MD simulations for (8-Br-)GTP at λ's in the interval $\langle 0.0,\lambda_{\max}\rangle$ with a λ increment of 0.05. For each λ two MD simulations of 1 ns were performed, one starting from an anti and the second from a syn conformation. Configurations of trajectories were stored every ps leading to 1000 frames for each trajectory. Configurations have been separated into sets according their conformational state ($\{q^{\text{anti}}_{\lambda}\}$, $\{q^{\text{syn}}_{\lambda}\}$).

The distributions for the switching energy differences $\{\Delta^{c_k,c_l}_{\lambda_i,\lambda_j}\}$ as well as for the switching probabilities $\{w^{c_k,c_l}_{\lambda_i,\lambda_j}\}$ and their cumulative values were obtained as described in Sec. II D 3 and schematically outlined in Fig. 5. Representative

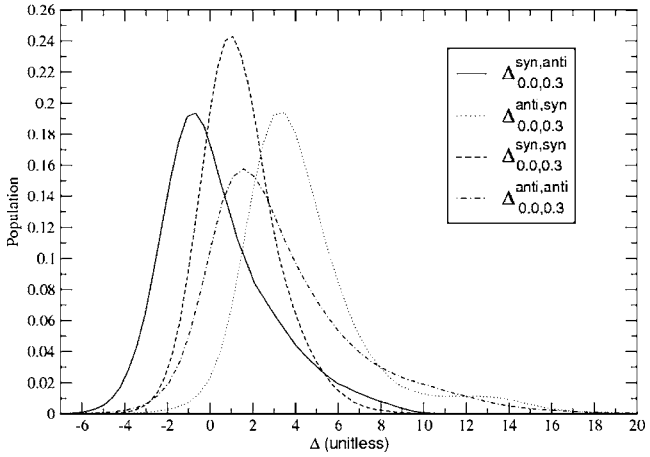

FIG. 8. Probability distributions of switching energy differences for GTP: $\Delta^{c_i,c_j}_{0.0,0.3}=\beta[E_{0.3}(q^{c_i}_{0.0})_r-E_{0.0}(q^{c_i}_{0.0})_r+E_{0.0}(q^{c_j}_{0.3})_s-E_{0.3}(q^{c_j}_{0.3})_s]$, where c_i, c_j are different conformational states: syn, anti and r, s are different configurations from MD trajectories.

examples of the switching energy differences and switching probability distributions are given in Figs. 8 and 9.

C. REMD mimicking

REMD mimicking as described in the Sec. II D 1 produces as output the time evolutions of conformational states (anti and syn) at all λ values that are included. Because of its speed we can run mimicked REMD for a very long time and thus obtain converged populations of anti and syn conformations for GTP and 8-Br-GTP. Typically, we simulate 500 000 cycles of mimicked REMD corresponding to 10^6 t^{elem} in real REMD (~2.5 μs).

Syn populations at individual λ values of mimicked REMD for GTP (500 000 cycles) with two replicas at $\lambda_{\max}=0.45$ combined with no or a single middle λ value are shown in Table I. Theoretically, syn populations at λ=0.0 should be the same for all mimicked REMDs. We do not suffer here from insufficient lengths of REMD simulations, but inaccuracies rather stem from the probability distributions of switching probabilities. Discrepancies in the syn

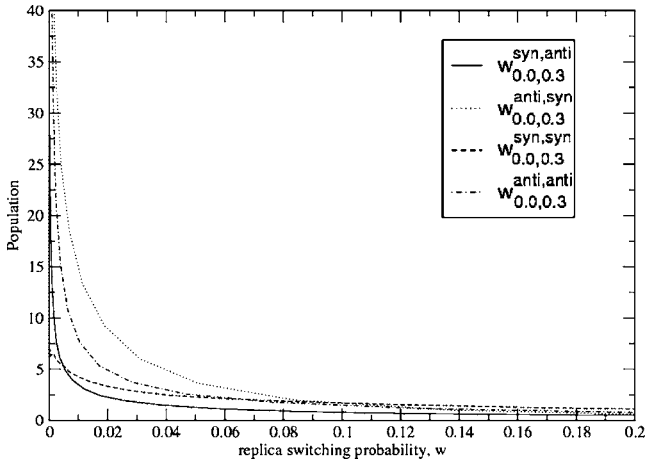

FIG. 9. Probability distribution of replica exchange switching probabilities for GTP: $w^{\text{syn,anti}}_{0.0,0.3}=\min[1,\exp(-\Delta^{\text{syn,anti}}_{0.0,0.3})]$ for GTP. Notice that probability $w^{\text{syn,anti}}_{0.0,0.3}$ is much lower than for the opposite replica switch $w^{\text{anti,syn}}_{0.0,0.3}$.

TABLE I. Syn populations at λ_0, $^1\lambda_1$, $\lambda_{max,0}$, $\lambda_{max,1}$ depending on the placement of $^1\lambda_1$ from mimicked REMD of GTP with two replicas at $\lambda_{max}=0.45$ over 500 000 cycles.

$^1\lambda_1$	None	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
[syn](λ_0)	0.053	0.047	0.053	0.050	0.044	0.043	0.049	0.047	0.081
[syn]($^1\lambda_1$)	···	0.058	0.095	0.133	0.163	0.228	0.287	0.295	0.298
[syn]($\lambda_{max,0}$)	0.302	0.303	0.296	0.301	0.305	0.307	0.299	0.306	0.302
[syn]($\lambda_{max,1}$)	0.300	0.301	0.295	0.299	0.303	0.304	0.298	0.304	0.300

populations at $\lambda=0.0$ can be explained from an insufficient sampling during the 1 ns MD simulation from which the probabilities are obtained. For λ values that are evident outliers compared to the other values, this simulation should probably be extended.

Largest deviations are expected for REMDs involving middle λ values, for which we multiply very small and very high probabilities (i.e., $^1\lambda_1=0.05$ and $^1\lambda_1=0.4$).

The root-mean-square deviation of [syn] at λ_0 divided by its average value when excluding the values obtained with $^1\lambda_1=0.05$ and $^1\lambda_1=0.4$ is calculated to be 0.068 indicating the relative inaccuracy in the distributions of switching probabilities.

The average of syn populations at λ_0 for runs with $^1\lambda_1$ between 0.1 and 0.35 amounts to $[\text{syn}]_{aver}=0.048$, which corresponds to a free energy difference $\Delta G^{GTP}_{syn\text{-}anti}=7.40$ kJ mol^{-1} at 298 K. As shown in Ref. 18 the obtained population of syn conformations from real *H*-REMD was 0.044 for GTP corresponding to a value of $\Delta G^{GTP}_{syn\text{-}anti}=7.63$ kJ mol^{-1}. 500 000 cycles of mimicking REMD using the same settings as the real REMD reported there ($\lambda=0.0;0.25,0.45,0.45,0.45,0.45$) gives [syn]=0.043, corresponding to $\Delta G^{GTP}_{syn\text{-}anti}=7.69$ kJ mol^{-1}. It shows that if most of the transitions can indeed occur only at λ_{max}, then REMD mimicking can produce very reasonable values of populations of the individual conformational states.

Therefore, REMD mimicking is not only useful as a tool for the optimization of REMD settings but can also as be used in itself for calculating the populations of conformations. We can see an analogy between our method and a free energy method that makes use of alternative pathways to calculate free energy differences more efficiently along unphysical thermodynamic cycles.[26] We want to note, however, that mimicked REMD covers very many of such pathways simultaneously.

D. Finding the optimal settings for REMD

In the previous paragraph we demonstrated how long mimicked REMD can approximate syn and anti populations for GTP and 8-Br-GTP. These simulations would correspond to several μs of real REMD. In real REMD, however, we can afford to simulate only limited time lengths (~tens ns), therefore it is crucial to find settings which corresponds to the highest possible REMD efficiency. We have used mimicked REMD, because it allows us to attempt REMD for hundreds of different combinations of λ's. From the time sequences of conformational states at different λ values we can choose the combination of λ's that produces the highest number of global conformational transitions, N_{gct}, which will ensure the fastest convergence of populations at λ_0.

1. GTP

Mimicked REMD (500 000 cycles) without middle replica and two replicas at $\lambda_{max}=0.45$ gives 5221 global conformational transitions, which is 5221/3=1740 global conformational transitions per CPU. The results for mimicked REMD with an increasing number of replicas at $\lambda_{max}=0.45$ are summarized in the Table II. It shows that the highest efficiency per CPU (7506/4=1877) is obtained for three replicas at λ_{max}.

In the next step we performed mimicked REMD with different numbers of replicas at $\lambda_{max}=0.45$ and one middle λ-value $^1\lambda_1\in\{0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4\}$ for 500 000 cycles [Fig. 10(a)]. With two replicas at $t^{total}_{\lambda_{max}}=10t^{elem}$ the highest values of N_{gct} are obtained for $^1\lambda_1=0.15$ (6069), $^1\lambda_1=0.2$ (6071), and for $^1\lambda_1=0.25$ (6089), which are higher than the values of N_{gct} obtained from REMD without middle λ having the same two replicas at $\lambda_{max}=0.45$. It means that putting one middle $^1\lambda_1$ can increase the efficiency of diffusion between $\lambda_0=0.0$ and $\lambda_{max}=0.45$. The efficiency per CPU for mimicked REMD containing one middle $^1\lambda_1=0.25$, however, is lower (1522 compared to 1877). Still the efficiency per CPU may be higher for REMD containing one middle lambda in combination with a higher number of replicas at λ_{max}. Figure 10 presents the efficiency of REMD containing one middle lambda with an increasing number of replicas at t^{elem}. It reveals that per CPU the most efficient set of λ's is [0.0, 0.25, 0.45, 0.45, 0.45, 0.45] with 12646/6=2108 global conformational transitions per CPU. It also shows that the differences due to the placement of $^1\lambda_1$ are more pronounced for settings with a higher number of replicas at λ_{max}.

Because the optimal REMD setting containing one $^1\lambda_1$ is more efficient per CPU than a set of λ's without any middle λ value, we continued to test REMD containing two middle λ values. As is explained in the methods Sec. II D 4 the optimal setting of REMD with two middle λ's will still have $t+t^{elem}$ replicas at $\lambda_{max}=0.45$, because this is the optimal

TABLE II. Number of global conformational transitions, N_{gct}, during mimicked REMD of GTP with $n=2,3,4,5$ replicas at $\lambda_{max}=0.45$ and no middle λ value after 500 000 cycles.

n replicas at $\lambda_{max}=0.45$	2	3	4	5
N_{gct}	5221	7506	8954	9911
N_{gct}/CPU	1740	**1877**	1791	1652

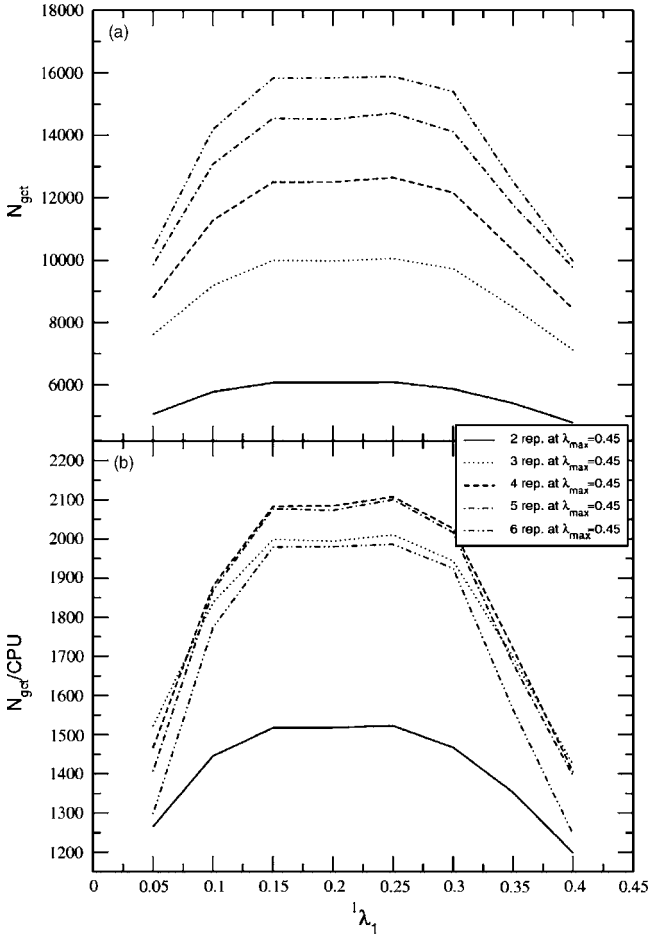

FIG. 10. Number of global conformational transitions, N_{gct} observed in 500 000 cycles of mimicked REMD of GTP with one middle replica $^1\lambda_1$ at different positions and with different numbers of replicas at λ_{max}=0.45: (a) absolute value and (b) value per CPU.

setting of REMD with one middle λ value. Table III shows N_{gct} per CPU for $n=4,5,6$ and two additional values $^2\lambda_1$ and $^2\lambda_2$. It shows that the optimal setting for REMD in this case is [0.0, 0.15, 0.3, 0.45, 0.45, 0.45, 0.45, 0.45]. However, its efficiency is slightly lower than for the optimal setting using only one middle $^1\lambda_1$ (2092 compared 2108 of global conformational transitions per CPU).

In conclusion the most effective setting predicted by mimicked REMD is [0.0, 0.25, 0.45, 0.45, 0.45, 0.45] with 12646/6=2108 global conformational transitions per CPU. This number corresponds to 500 000 REMD mimicking cycles or 10^6 elementary switching periods of real REMD. This means that one can expect ~13 global conformational transitions per 1000 t^{elem} of real REMD, which amounts to 2.5 ns with t^{elem}=2.5 ps. From an extrapolation of the linear increase of N_{gct} as a function of time, over ten individual 5000 cycles mimicked REMD simulations we estimate the REMD equilibration period to be about 200 t^{elem} (0.5 ns in real REMD). For comparison, in real REMD we reported that N_{gct} increases by ~17 N_{gct} per 1000 t^{elem} and that a REMD equilibration of 249 t^{elem} was required.[18]

2. 8-Br-GTP

8-Br-GTP has a higher conformational barrier between the anti and syn conformations than GTP, therefore we have found a higher value of λ_{max}=0.7 and the optimal REMD will probably require more middle replicas. For this reason we will not use a systematic search for the optimal settings as for GTP, but rather use a more efficient approach as described in Sec. II D 4.

We initiate the optimization of the middle λ's using five replicas at λ_{max}=0.7 and obtain $^1\lambda_1^{opt}$=0.5 as the optimal value if only one middle value of λ is taken into account, with N_{gct}=1079 per CPU (Table IV). When increasing the number of middle λ values to two, we restricted the searching ranges to $^2\lambda_1 \in \langle 0.05, 0.5\rangle$ and $^2\lambda_2 \in \langle 0.5, 0.65\rangle$. Similarly the optimizations were extended to three and four middle λ values. The N_{gct} per CPU for different numbers of middle λ values in combination with five replicas at λ_{max} are summarized in Table IV, where arrows indicate the steps in the optimization. As can be seen from this table, the optimal setting containing four middle λ values in combination with five replicas at λ_{max} is less efficient (1252 global conformational transitions per CPU) than the optimal three middle λ-values setting (1274). Because this may still be caused by too few conformational transitions at λ_{max}, the same optimal settings of four middle λ values [0.2; 0.4; 0.55; 0.65] was tested in combination with $n=6,7,8$ and indeed for $n^{opt}(m=4)=7$ we obtained a more efficient setting (1315 N_{gct} per CPU).

Taking into account the inequalities: $n^{opt}(m>4) \geq n^{opt}(m=4)=7$ and $n^{opt}(m<4) \leq n^{opt}(m=4)=7$ that are described in Sec. II D 4, the optimization procedure was continued by including five middle λ values in combination with $n=7$ and subsequently with $n=8$. It appears that $n^{opt}(m=5)=7$, but that its efficiency is lower (1305) than for the optimal four middle λ-values setting (1315 N_{gct} per CPU), meaning that the overall optimal scheme has ≤4 middle λ values with $n \leq 7$. Because the optimal three middle λ-values setting $[n^{opt}(m=3)=6]$ is also not giving a higher efficiency, the overall most efficient setting for 8-Br-GTP has thus been determined as $^4\lambda_i^{opt} \in [0.0;0.2;0.4;0.55;0.65;0.7;0.7;0.7;0.7;0.7;0.7;0.7]$.

N_{gct} per CPU (1315) is lower for 8-Br-GTP than for GTP (2108) indicating that REMD of 8-Br-GTP will be computationally more demanding. Using the optimal set for 8-Br-GTP of 12 replicas we estimate ~16 global conformational transitions during 1000 t^{elem} of real REMD (~2.5 ns) and the equilibration period to be ~180 t^{elem} (based on a linear extrapolation of time dependence of N_{gct} as function of time from ten individual runs of 5000 mimicked REMD cycles).

IV. DISCUSSION

The present study on the GTP/8-Br-GTP two state model systems deepens our understanding about REMD efficiency. For simplicity, we assumed that conformational transitions occur only at the highest lambda, λ_{max} in the case of *H*-REMD, or T_{max} for *T*-REMD. To obtain a different conformational state for a given replica at λ_0, the replica has to diffuse from λ_0 through the middle λ values to λ_{max}, where a conformational change has to occur. The system subsequently needs to diffuse from λ_{max} back to λ_0. Such process we named a global conformational transition. The advantage

TABLE III. Number of global conformational transitions, N_{gct} per CPU for mimicked REMD of GTP using $n=4,5,6$ replicas at $\lambda_{max}=0.45$ and two additional λ-values ${}^2\lambda_1$, ${}^2\lambda_2$. The maximum is obtained for ${}^2\lambda_1=0.15$ and ${}^2\lambda_2=0.3$, at $n=5$.

$n=4$	${}^2\lambda_1$				${}^2\lambda_2$			
		0.1	0.15	0.2	0.25	0.3	0.35	0.4
	0.05	1614	1741	1726	1727	1697	1413	1269
	0.1	⋯	1787	1807	1879	1964	1692	1622
	0.15	⋯	⋯	1802	1898	**1990**	1807	1707
	0.2	⋯	⋯	⋯	1839	1988	1767	1761
	0.25	⋯	⋯	⋯	⋯	1841	1701	1719
	0.3	⋯	⋯	⋯	⋯	⋯	1678	1688
	0.35	⋯	⋯	⋯	⋯	⋯	⋯	1484
$n=5$	${}^2\lambda_1$				${}^2\lambda_2$			
		0.1	0.15	0.2	0.25	0.3	0.35	0.4
	0.05	1626	1814	1755	1766	1746	1446	1261
	0.1	⋯	1834	1871	1960	2030	1795	1644
	0.15	⋯	⋯	1813	1948	**2092**	1867	1803
	0.2	⋯	⋯	⋯	1945	2070	1845	1790
	0.25	⋯	⋯	⋯	⋯	1931	1776	1795
	0.3	⋯	⋯	⋯	⋯	⋯	1742	1740
	0.35	⋯	⋯	⋯	⋯	⋯	⋯	1470
$n=6$	${}^2\lambda_1$				${}^2\lambda_2$			
		0.1	0.15	0.2	0.25	0.3	0.35	0.4
	0.05	1562	1742	1742	1709	1702	1353	1167
	0.1	⋯	1781	1827	1920	2026	1722	1524
	0.15	⋯	⋯	1812	1917	**2075**	1830	1715
	0.2	⋯	⋯	⋯	1947	2069	1795	1648
	0.25	⋯	⋯	⋯	⋯	1906	1739	1714
	0.3	⋯	⋯	⋯	⋯	⋯	1715	1716
	0.35	⋯	⋯	⋯	⋯	⋯	⋯	1430

of monitoring the N_{gct} compared to the number of roundtrips between the lowest and highest temperature as proposed by Trebst *et al.*[14–16] comes from the fact that many roundtrips do not necessarily ensure any conformational changes at λ_0. For example for GTP, we have observed simulations in which there were many roundtrips thanks to relatively high values of $w^{anti,anti}_{\lambda_i,\lambda_j}$, while $w^{anti,syn}_{\lambda_i,\lambda_j}$ were very low. In such cases a syn conformation that occurs at λ_{max} will not be able to efficiently diffuse down to λ_0. Another advantage of counting N_{gct} is that if $P^{transit}_{\lambda\neq\lambda_{max}}(t^{elem})\neq 0$ it does not require a replica to diffuse all the way up to λ_{max} because a conformational transition within real REMD can already occur at any λ value. The disadvantage of monitoring N_{gct} is that defined stable conformational states are required.

We described the approach for finding the optimal λ_{max} at which conformational transitions occur in sufficiently short time, which is, however, typically still much longer than the elementary switching period, t^{elem}, between switching attempts. The occurrence of conformational transitions shows a sigmoidal time dependency, from which a relatively long $t^{transit}_{\lambda_{max}}$, $t^{total}_{\lambda_{max}}$ can be estimated. The presented approach using n multiple replicas at λ_{max} allows several replicas to

TABLE IV. Number of global transitions, N_{gct} per CPU for the mimicked REMD of 8-Br-GTP, having optimal set of m middle lambdas ${}^m\lambda_i^{opt}$ combined with n replicas at $\lambda_{max}=0.7$. The maximum is obtained for $m=4$ and $n=7$. The arrows indicate the sequence of steps during the optimization procedure.

m	0		1		2		3		4		5
Set of ${}^m\lambda_i^{opt}$	None		[0.5]		[0.4,0.55]		[0.35,0.5,0.65]		[0.2,0.4,0.55,0.65]		[0.2,0.4,0.45,0.55,0.65]
N_{gct}/CPU, $n=5$	159	→	1079	→	1263	→	1274	→	1252		
									↓		
N_{gct}/CPU, $n=6$							1301		1264		
							↑		↓		
N_{gct}/CPU, $n=7$							1297	←	1313	→	1305
									↓		↓
N_{gct}/CPU, $n=8$									1282		1266

spend the required time $t_{\lambda_{max}}^{total}$ at λ_{max} (in parallel). At the same time it allows for frequent switching attempts between $\lambda_0, \ldots, \lambda_i, \ldots \lambda_{max,0}$. In order to optimize the number of replicas at λ_{max} as well as the number and placement of λ values between λ_0 and λ_{max}, we present a REMD mimicking approach. As this approach propagates the conformations based on calculated probabilities, it is very fast and allows for simulations that correspond to μs time scales in real REMD.

This approach was inspired by the following analogy with real REMD. In real REMD "parallel" MD simulations are halted from time to time and switch their temperatures or Hamiltonians depending on the current energies. If we collect the switching probabilities into probability distributions then we can mimic the REMD simulation by sampling a switching probability from the probability distribution and subsequently perform the switch depending on this probability. Although real REMD switching probabilities converge to the same distribution, it builds up this distribution relatively inefficiently because for every structure at a particular time point, only one partner conformation is selected to attempt a replica switch. In our approach, all possible pairs of structures at a given set of λ_i, λ_j are used to estimate the distribution of switching probabilities. In many cases quite long REMD simulations are required to obtain a statistically sound probability distribution of switching probabilities. In this work we perform relatively short MD simulations at different λ's (*H*-REMD) or temperatures (*T*-REMD) starting from different conformational states and subsequently calculate the corresponding distributions of energy differences using different Hamiltonians or temperatures. These distributions are then used to calculate the probability distributions of switching probabilities. Together with estimates of the probability of conformational transitions at λ_{max} it allows us to mimic several μs of REMD simulations in a few minutes. We can then study the effect of different middle λ sets on the REMD efficiency, characterized by N_{gct} per CPU. For example, we clearly showed that increasing the number of replicas does not necessary increase the efficiency (per CPU) of REMD simulations. This means that "blind" brute force applications of REMD may be very inefficient.

A systematic search for the optimal λ settings that give the highest value of N_{gct} per CPU is affordable for simpler systems with a small number of replicas. However, the optimization process for complex systems requiring many replicas this can lead to a huge number of combinations. For such cases set we have suggested approaches to reduce the number of relevant combinations. The whole proposed optimization scheme is largely automated and fully parallelized.

Once the optimal settings have been found, REMD mimicking can make two more practical predictions for real REMD: it can estimate: (1) the equilibration time and (2) the required length of an REMD simulation to reach a given number of N_{gct}. For GTP this later time estimate showed an accuracy of about 20% in real REMD. This allows the user to estimate the feasibility of the real REMD simulation and to allocate the needed computational time in advance.

The most important advantage of the presented optimization scheme for REMD settings compared to the feedback-optimized parallel tempering approach[14–16] is its speed. While feedback optimization of the 36-residue villin headpiece subdomain HP-36 requires several years of CPU time[14–16] optimization by REMD mimicking can be performed within one week. This is a crucial factor for REMD simulations of complex system such as proteins. On the other hand, our approach requires some preliminary knowledge of stable conformational states, at least of the most dominant ones. Starting with an incomplete set of stable conformational states can lead to not completely optimal REMD settings. However, real REMD using partially optimized settings may suffice to reveal additional stable conformational states, which can then be used to refine the REMD settings. The whole process can be repeated iteratively until the real REMD leads to converged populations of the individual conformations. Note that one set of REMD generated conformations may yield part of the simulations that are used to calculate the probability distributions required for refining the λ settings.

We also showed that REMD mimicking is not only useful for the optimization of REMD settings, but also by itself can give us reasonable estimates of conformational populations. For GTP we obtained an excellent agreement for the syn population at $\lambda_0=0.0$ from mimicked REMD ($[syn]_{aver}=4.8\%$), as compared to real REMD ($[syn]=4.4\%$) presented in our other work.[18] The advantage of mimicked REMD compared to thermodynamic cycle approaches is the fact that global conformational transition can occur through many different pathways, which are directly counted in a massively parallel manner in REMD mimicking using probability distributions of all combinations of switching probabilities. Usage of probability distributions, instead of average values takes into account the conformational variety of structures belonging to the same stable conformational region.

V. CONCLUSIONS

We have presented an algorithm that mimics replica exchange (REMD) simulations by a stochastic propagation in time of conformational states rather than explicit MD simulations. The approach was demonstrated for Hamiltonian REMD simulations on two model systems, GTP and 8-Br-GTP, for which two stable conformations are known, but the potential energy barrier separating them is too high to be crossed in normal MD. The method is, however, also readily applicable to temperature REMD.

We have shown that mimicked REMD can be used to serve three different purposes: (1) it can be used to obtain the optimal set of λ values by allowing an extremely fast attempt to try different REMD settings; (2) it gives the user an estimate of the simulation length in real REMD simulations, both for the equilibration of conformational states over the replicas and for the time required to obtain reasonably converged populations; and (3) it can make estimates of such populations directly which were shown to match remarkably well with real REMD simulations.

We are convinced that our method can contribute significantly to deepen our understanding of the processes governing REMD simulations. For many different applications, it

will help to design more efficient REMD simulations for systems as are described in this work, but also for many more complex systems.

ACKNOWLEDGMENTS

This work was supported by the Netherlands Genomics Initiative in the context of Horizon Breakthrough Grant No. 050-71-043 (postdoc stay of J.H.).

[1] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[2] M. C. Tesi, E. J. J. van Rensburg, E. Orlandini, and S. G. Whittington, J. Stat. Phys. **82**, 155 (1996).
[3] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[4] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[5] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[6] Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).
[7] D. J. Earl and M. W. Deem, J. Phys. Chem. B **108**, 6844 (2004).
[8] D. A. Kofke, J. Chem. Phys. **117**, 6911 (2002).
[9] D. A. Kofke, J. Chem. Phys. **121**, 1167 (2004).
[10] A. Kone and D. A. Kofke, J. Chem. Phys. **122**, 206101 (2005).
[11] C. Predescu, M. Predescu, and C. Ciobanu, J. Chem. Phys. **120**, 4119 (2004).
[12] C. Predescu, M. Predescu, and C. Ciobanu, J. Phys. Chem. B **109**, 4189 (2005).
[13] N. Rathore, M. Chopra, and J. J. de Pablo, J. Chem. Phys. **122**, 024111 (2005).
[14] S. Trebst, D. A. Huse, and M. Troyer, Phys. Rev. E **70**, 046701 (2004).
[15] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, J. Stat. Mech.: Theory Exp. **2006**, P03018 (2006).
[16] S. Trebst, M. Troyer, and U. H. E. Hansmann, J. Chem. Phys. **124**, 174903 (2006).
[17] T. Läppchen, A. F. Hartog, V. A. Pinas, G. J. Koomen, and T. den Blaauwen, Biochemistry **44**, 7879 (2005).
[18] J. Hritz and C. Oostenbrink (in preparation).
[19] M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Burgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Krautler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. Van Gunsteren, J. Comput. Chem. **26**, 1719 (2005).
[20] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[21] R. W. Hockney, Methods Comput. Phys. **9**, 136 (1970).
[22] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[23] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, J. Chem. Phys. **102**, 5451 (1995).
[24] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren, J. Comput. Chem. **25**, 1656 (2004).
[25] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, Chem. Phys. Lett. **222**, 529 (1994).
[26] A. E. Mark, W. F. Van Gunsteren, and H. J. C. Berendsen, J. Chem. Phys. **94**, 3808 (1991).